An Empirical Mode Decomposition Fuzzy Forecast Model for Air Quality

Bo-Lun Chen
Huaiyin Institute of Technology

Guo-Chang Zhu (zhuguochang1996@163.com)
Huaiyin Institute of Technology

Yi-Yun Sheng
Huaiyin Institute of Technology

Qian Xie
Huaiyin Institute of Technology

Min Ji
Huaiyin Institute of Technology

Yong-Tao Yu
Huaiyin Institute of Technology

Research Article

Keywords: Air quality, health, air pollution, adaptive fuzzy inference system, pollutants

Posted Date: July 9th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-676389/v1

An empirical mode decomposition fuzzy forecast model for air quality

Bo-Lun Chen[1], Guo-Chang Zhu[1,*], Yi-Yun Sheng[1], Qian Xie[1], Min Ji[1], and Yong-Tao Yu[1]

[1]Faculty of Computer and software Engineering, Huaiyin Institute of Technology, Huaian, Jiangsu, 223003,China
[*]zhuguochang1996@163.com

ABSTRACT

Air quality is related to people's health. Severe air pollution can cause respiratory diseases, while good air quality is beneficial to physical and mental health. Therefore, the prediction of air quality is very important. As an important algorithm for signal analysis, empirical mode decomposition can analyze the change trend of air quality well, smooth the complex and changeable air quality data, and get the change trend of air quality under different time scales. According to the change trend under different time scales, the extreme learning machine is used for training, and the corresponding prediction value is obtained. The adaptive fuzzy inference system is used for fitting to obtain the final air quality prediction result. The experimental results show that the signal decomposition fuzzy prediction model has a good learning ability and has good accuracy in predicting the concentration of various pollutants in air quality.

Introduction

Medical research shows that air has a direct effect on the blood circulation, immunity, cardiovascular system, and nervous system of the human body. Breathing fresh air is good for people's health. Inhaling dirty air for a long time can cause many physical and psychological problems, such as dizziness, chest tightness, difficulty in concentrating, and even induce some diseases. At the same time, severe air pollution can also greatly reduce visibility, make it difficult for people to travel, and even cause traffic accidents. Air pollution also has an impact on the ecological environment. It can change the diversity and stability of the natural ecosystem structure. In addition, severe air pollution can cause physical damage to the building, affecting the appearance of the building and the service life.

Since the first industrial revolution, industrialization and urbanization use a large number of resources, resulting in increased exhaust emissions and plummeting air quality, which brings unprecedented negative effects on people's living environment. A major air pollution accidents, for governments to understand the air quality is very important. At the same time, there are more and more researches on air quality in academic circles. And the prediction of air quality index is undoubtedly an important link among them.

There are many prediction methods for air quality. The early prediction method in the world is potential prediction. However, due to its only consideration of meteorological factors, the prediction results are not accurate, so it does not belong to the current mainstream prediction methods. Numerical prediction and statistical prediction are the most popular methods for air quality prediction in the world. Numerical prediction is a method based on physics and combined with knowledge of many subjects. It has good accuracy, timeliness and efficiency. However, this method is very difficult, needing accurate data support. The problem, such as large amount of calculation, makes it only suitable for large institutions or government agencies. Different from numerical forecasting, statistical forecasting does not need to understand the development law of air quality and the relationship between various factors. It is a method that only needs to analyze the law and make prediction through data. It is suitable for institutions, governments and individuals. With its obvious advantages, statistical prediction has received more and more attentions in the research of air quality prediction. Common statistical forecasting methods include grey model, clustering, and neural networks.

By analyzing and determining the relationship between the concentration of air pollutants and environmental factors, clustering can get the correlation between each element, so as to predict the air quality. The concentration of air pollutants is characterized by nonlinear and non-stationary characteristics, including high noise levels and outliers. To solve this problem, Chen et al. took PM2.5 as an example, decomposed PM2.5 data by ensemble empirical mode decomposition, selected the most effective set of data and projected it by least square method[1]. Clustering can not only measure the factors that affect the concentration of air pollutants, but also improve the prediction accuracy and reduce the time complexity. Lee et al. clustered the spatial heterogeneity of PM2.5 concentration and used a mixed effect model to make aerosol optical thickness PM2.5 concentrations of robust predictors, and through the remote sensing map to predict the concentration of PM2.5[2]. Mahajan et al.

clustered the geographic distances of 557 urban stations in Taiwan and used a mixed model to make predictions. The experiment shows that clustering based on geographical distance can reduce the prediction error rate and also reduce the calculation time[3]. Xiao et al. used clustering to divide the Chinese space into seven regions. Each region is trained by random forest, generalized additive model and limit gradient lifting, and finally the additive set model is used to combine the prediction[4].

Grey model can be processed according to less time series data and extract effective information from known information. It is a first-order model with only a single variable and its prediction curve is relatively simple. Wu et al. proposed a grey Holt-Winters model, which used the grey cumulative generation to deal with the unstable air quality data and the Holt-Winters method to deal with the seasonal effects, with high prediction accuracy[5]. Xiong et al. Proposed the nonlinear multivariable GM(1,N) model based on interval grey number sequence. It deduces the upper and lower limits of the interval gray number according to the kernel formula and gray radius to simulate and predict[6]. With the less amount of data, the gray model can predict very well. But with a large amount of data, the effect of grey model is not ideal. Other methods of machine learning can better solve this problem. Zakeviciute et al. used traffic vehicle flow data to predict air quality data through the decision tree[7]. The experiment shows that the prediction accuracy is improved when the traffic flow is large in the daytime, and it is suitable for the countries with low economy. Gu et al. proposed a heuristic recursive air quality forecast model[8]. It extracts air quality data and meteorological data, and makes cycle prediction with support vector machine. Castelli et al. improved the regression of the support vector machine by taking the radial basis function as the core of the support vector regression machine[9]. The improved support regression machine was used to predict air quality and achieved good results. Lyu et al. proposed a bias correction model, including feature selection, clustering, error estimation and interpolation[10]. The model uses the historical relationship between the predicted variables, the observed variables, and the model bias to make the current forecast. Liu et al. proposed an n-step recursive prediction model based on Seq2Seq to predict air quality[11]. It replaces the encoder in Seq2Seq with a fully connected encoder to speed up the training process. At the same time, combined with the position embedding technology, the fully connected encoder can discover the timing relationship between the sequences.

Deep learning is also widely used in air quality prediction due to its powerful nonlinear fitting ability. Ray et al. proposed a prediction model based on deep learning[12]. It uses correlation analysis to find the most appropriate feature set which is used to predict air quality. Qi et al. embedded feature selection and semi-supervised learning in different levels of the deep learning network, and integrated air quality interpolation, prediction and feature analysis into a model[13]. Yang et al. proposed a hybrid model combining complementary integrated empirical mode decomposition, improved cuckoo search, differential evolution algorithm and Elman neural network to predict air quality[14]. Cheng et al. used a recursive neural network to extract information from the input sequence and measure the impact of other sites, using the full connection layer to obtain the predicted results[15]. Yi et al. considered the spatial correlation of air pollutants, processed sparse air quality data with spatial transformation components, and fused the data affecting air quality of multiple cities with neural distribution structures[16]. Li et al. proposed an air quality prediction method based on spatiotemporal deep learning, which extracted inherent air quality characteristics with stacked autoencoder models and trained them in a greedy layer-by-layer manner. It can simultaneously predict air quality at multiple sites and shows temporal stability over the seasons[17]. Peng et al. proposed a learning method based on an extremum learning machine, which learns continuously as the data increases, combined with multiple linear regressions and multi-layer perceptron networks, to predict air quality[18]. Kabi et al. combined the confidence expert system with deep learning, and optimized the confidence expert system so as to find the nonlinear dependence among the relevant variables, so as to predict air quality[19]. Bai et al. proposed a seasonal stacked automatic encoder model combining seasonal analysis and deep feature learning[20]. Kendall correlation coefficient method was used to explore the internal relationship between air quality and seasons. Deep neural network was used to extract features and regression stacked automatic encoder was used to conduct prediction. Pardo et al. used Long Short Term Memory (LSTM) to predict air quality over the next 12 and 24 hours[21]. Athira et al. used a variety of deep learning methods to predict the air quality, including recursive neural networks, LSTM and gated recursive units[22]. Soh et al. combined artificial neural network, convolutional neural network and LSTM to extract spatio-temporal relations[23]. That is, the trend extracted from the correlation between the adjacent positions and the similar positions in the time horizon was applied to predict air quality. Zhou et al. proposed a deep multi-output model based on LSTM[24]. The model combines three deep learning algorithms, including discrete neurons, gradient descent and L2 regularization, to extract the key factors of spatio-temporal relationship and reduce the error accumulation and propagation in multi-step advanced air quality forecast. Lin et al. proposed an air quality prediction system based on neural networks[25]. Historical time series data were used to derive a set of fuzzy rules or neural fuzzy networks to predict future air pollutant concentrations and environmental factors. Yan et al. used the deep learning network model based on spatiotemporal clustering to establish a multi-time and multi-site prediction model for air quality in Beijing. Compared with back propagation, the model has a good prediction effect[26]. Wang et al. proposed a deep space-time integration model[27]. The model adopts the division strategy based on weather pattern for clustering, and generates spatial data of relative stations and relative regions by analyzing the Granger causality between stations, so as to find spatial correlation. Finally, the LSTM forecasts different types of data to understand the long-term and short-term correlation of air quality.

Most of the researches on air quality focus on the prediction of data from multiple stations, and fail to fully analyze the variation trend and internal relationship of air quality of a single city station. In this paper, the historical data of air quality of a single city is analyzed to explore its own internal laws, so as to predict air quality.

Results

Data Set

In this paper, the air quality of Huaian in Jiangsu province was selected for prediction. As shown in Figure 1, Huaian has several advantages: Huaian is located on the north-south boundary of China, it is affected by the monsoon climate. The climate change is obvious, the variation trend of air pollutant concentration is obvious, thus, it has a good representative; the economic development is good, with more and accurate air quality detection points; the terrain is mainly plain and part of hilly; There are many rivers in the territory and there is a large lake. The environment is changeable and there are a lot of accurate data, so it has a high research value. Therefore, the selection of Huaian air quality data for the experiment of the method can show whether the method has good learning ability and the prediction accuracy in the face of complex and changeable situations.

The data came from a data center of the China Ministry of Ecology and Environment(http://datacenter.mee.gov.cn),

Figure 1. Huaian relief map

collected in days, including CO, nitrogen dioxide (NO_2), ozone (O_3), PM10, PM2.5, sulfur dioxide (SO_2), and Air Quality Index (AQI). The data are shown in figure 2. The data are highly accurate and scientific, which is of great significance to the algorithm prediction experiment.

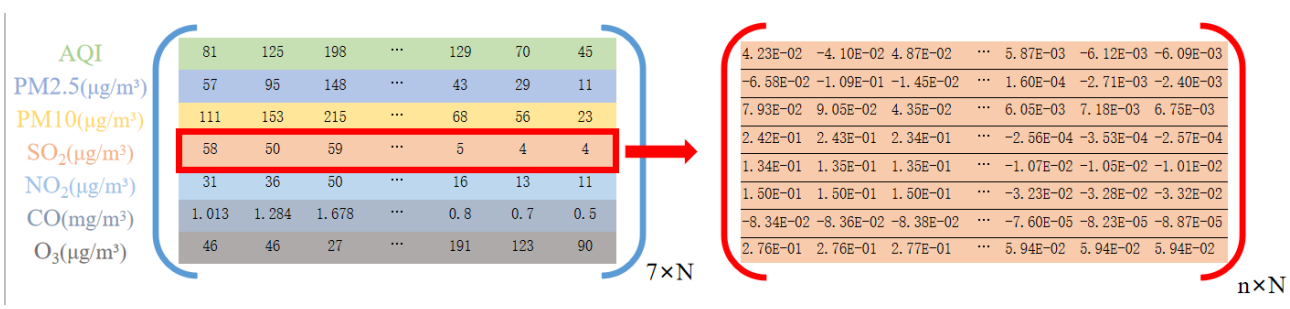

Figure 2. Data Form. The matrix on the left is 7 factors of the original data, with a total of N days; the matrix on the right is the concentration of an pollutant concentration (take SO_2 as an example) after normalization and decomposition.

Evaluation Standard

In order to evaluate the accuracy of the prediction, the following evaluation criteria are selected in this paper:

(1) Mean Square Error (MSE) is the most commonly used evaluation standard to test the Error between the predicted value and the true value. The formula is as follows:

$$MSE = \frac{\sum (y - y')^2}{N} \tag{1}$$

(2) Mean Absolute Error (MAE), namely the Mean of Absolute Error, is used to reflect the actual situation of the predicted value Error. The formula is as follows:

$$MAE = \frac{\sum |y - y'|}{N} \tag{2}$$

(3) Standard Deviation (SD), the arithmetic square Deviation of variance, which is used to reflect the dispersion between the predicted value or the true value and the mean value. Taking the true value as an example, the formula is as follows:

$$SD = \sqrt{\frac{\sum (y - \overline{y})^2}{N}} \tag{3}$$

Where y' represents the predicted value, y represents the true value, $\overline{y}$ represents the average value of the true value, and N represents the number of predicted or true values.

Contrast Experiment

Three groups of comparative experiments were selected for the experiment: ELM, Back Propagation (BP) neural network, and Nonlinear Auto Regressive (NAR) neural network. Their introductions are as follows:

(1) BP neural network is a multi-layer feedforward neural network[28], which is widely used in many research fields. The training process of the BP neural network is divided into two stages. First, data are propagated forward from the input layer to the hidden layer and then to the output layer. The error is then propagated back from the output layer to the hidden layer and then to the input layer.

(2) The NAR neural network[29] is a time-series neural network that uses itself as a regression variable. That is, a linear combination of random variables at certain moments to describe a nonlinear target moment. The NAR neural network is widely used in various forecasting models and has good results.

Experimental Results and Analysis

Decomposition Result

SO_2 concentration in Huaian has normalized and the EMD decomposition has carried out. The original data and decomposition results are shown in Figure 3. The first one is the original data and the rest are the IMF. As can be seen from the figure, the changes of the original data are complex and diverse without obvious rules, but they show a downward trend over time. It reflects the government's attention to the effect of air quality control and the public. Through multiple decomposition, the original unstable and non-smooth data gradually become smooth. As can be seen from the top five decomposition, after a period of stable sulfur, there is a period of relatively obvious fluctuations. This is due to the centralized treatment of air quality, failed to effectively continue, resulting in the recurrence of pollution. It can be seen that air quality control is a long-term process, and temporary control cannot completely eliminate pollution. It can be seen from the last three decomposition that the overall change trend is fluctuating, and the last one shows a downward trend, just like the overall trend of the original data. It can be also seen that the EMD decomposition is very effective for trend analysis of the data under different time scales.

Short-term Prediction

Short-term prediction was made for each index of air quality. That is, the data of the top six days were used to predict the value of the seventh day, and the prediction results are shown in Figure 4. As can be seen from the figure, BP neural network, ELM and NAR neural network have similar performances. These three neural networks only calculate the predicted value of air quality, and fail to fully explore the internal relationship of air quality, so the performances are close. ELM has the best effect and the error range of the three indicators is small. This is because ELM performs better when the data volume is small. The EMD-FPM has good accuracy and generalization performance in the prediction of air quality index or air pollutants, which is about 30% higher than other algorithms. All the three evaluation indexes are superior to other algorithms, which means that the EMD-FPM algorithm not only has high accuracy, but also has good adaptability to special situations.

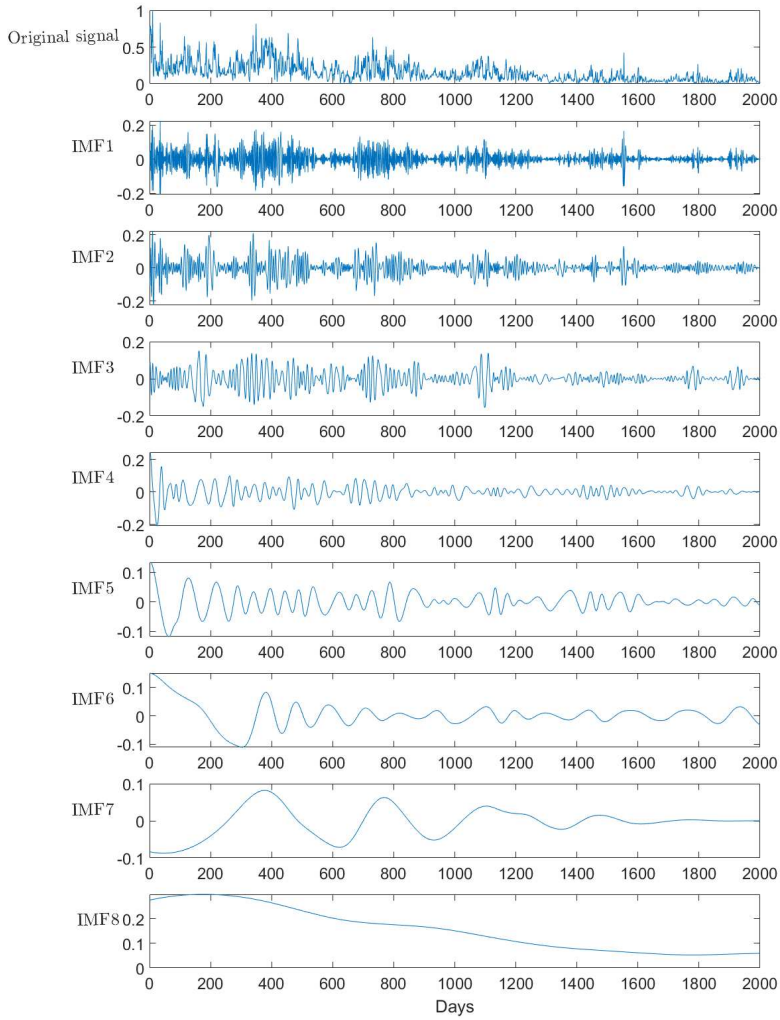

Figure 3. EMD:take sulfur dioxide as an example

Long-term Prediction

Long-term prediction was made for each indicator of air quality. That is, the data of the first 11 days were used to predict the value of the 12th day, and the prediction results are shown in Figure 5. As shown by the long-term prediction results, the EMD-FPM is far superior to other algorithms. The prediction accuracy is 30% 40% higher than other algorithms. Compared with the short-term prediction, the error of each evaluation index in all algorithms is increased, which is the inevitable result of the long-term prediction. However, the error increase of the EMD-FPM can be ignored, while other algorithms grow significantly. Therefore, the EMD-FPM also has good performance in long-term prediction.

Discussion

Empirical Mode Decomposition

Empirical Mode Decomposition (EMD) is an adaptive signal analysis method proposed by Huang et al.[30]. Different from the general signal analysis method, it does not need to set the basis function. Instead, by analyzing the trend of the signal itself in the characteristic time scale, the original signal is decomposed into sub-signals of different frequencies. This is a method with a wide range of applications and suitable for all kinds of signal analysis. The EMD first smoothes the signal and then decomposes the trend of different characteristics in the signal at time scales. The decomposed sequence of different characteristic time scales is called Intrinsic Mode Function (IMF). Figure 6 is a schematic diagram of the EMD.

The EMD is to find the extreme point of the original signal, where the maximum is fitted to form the upper envelope, and the minimum is fitted to form the lower envelope, and then the mean values of the two envelopes are taken to form the mean envelope.

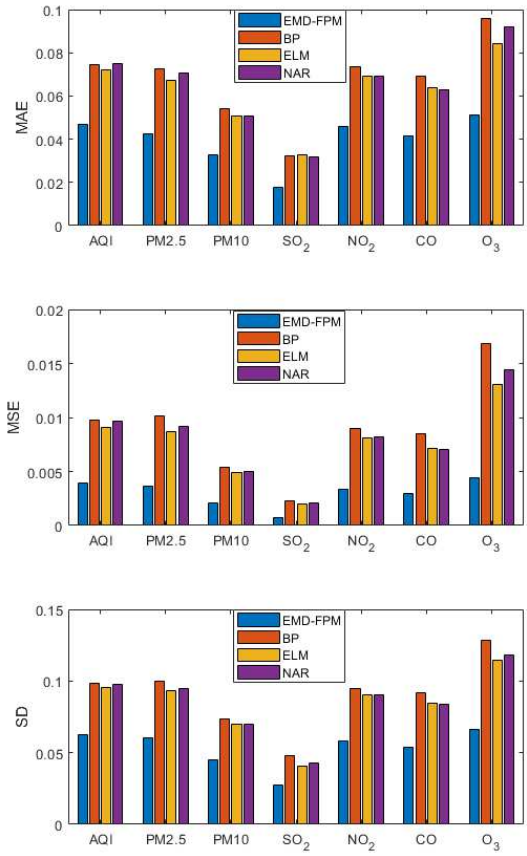

Figure 4. Short-term forecast comparison

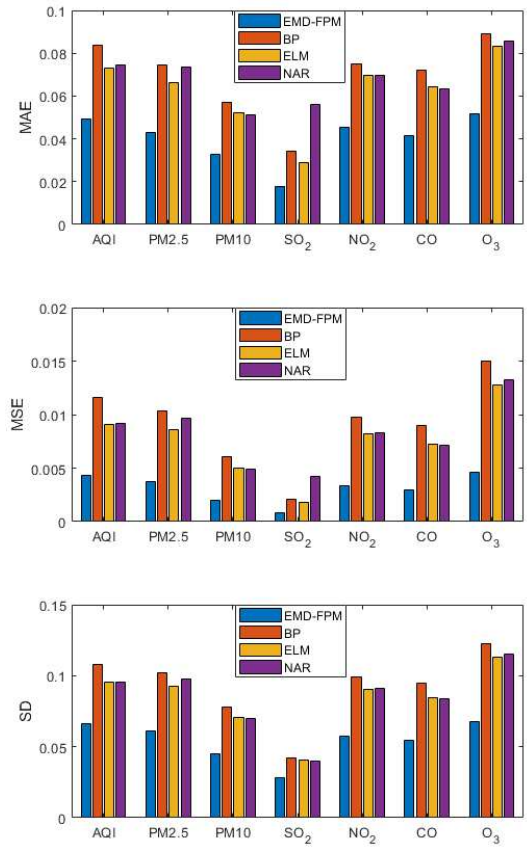

Figure 5. Long-term forecast comparison

Extreme Learning Machine

Extreme Learning Machine (ELM) is an improved single hidden layer feed forward neural network proposed by Huang Guangbin[31] in 2004. Figure 7 is the schematic diagram of the ELM. Different from other neural networks, the ELM training process is extremely simple. The number of hidden nodes can be set and the only optimal solution can be obtained after one training. Compared with the traditional gradient algorithm, the ELM has the advantage of extremely fast learning speed and solving the problems such as over fitting and local minimum. Meanwhile, its generalization ability performs better in some applications.

Suppose the excitation function of an ELM is $g(x)$, and there are N samples and L hidden layer nodes. The input is X, the expected output is T, and β is the output weight of the hidden layer. For a set of inputs x_i, given the corresponding expected outputs t_i, input layer nodes to hidden layer input weight a_i, bias b_i, and hidden layer output weight β_i, then:

$$\sum_{i=1}^{L} g(x_i a_i + b_i)\beta_i = t_i \tag{4}$$

H is the output matrix of the hidden layer, then:

$$H = \begin{bmatrix} g(x_1 a_1 + b_1) \\ \cdots \\ g(x_N a_N + b_N) \end{bmatrix}_{N*L} \tag{5}$$

The output of the ELM can be denoted as the following matrix:

$$H\beta = T \tag{6}$$

Since a_i and b_i in the ELM are randomly determined, H has also been determined, so the training process of the ELM is actually the process of solving β, then:

$$\beta = H^{-}T \tag{7}$$

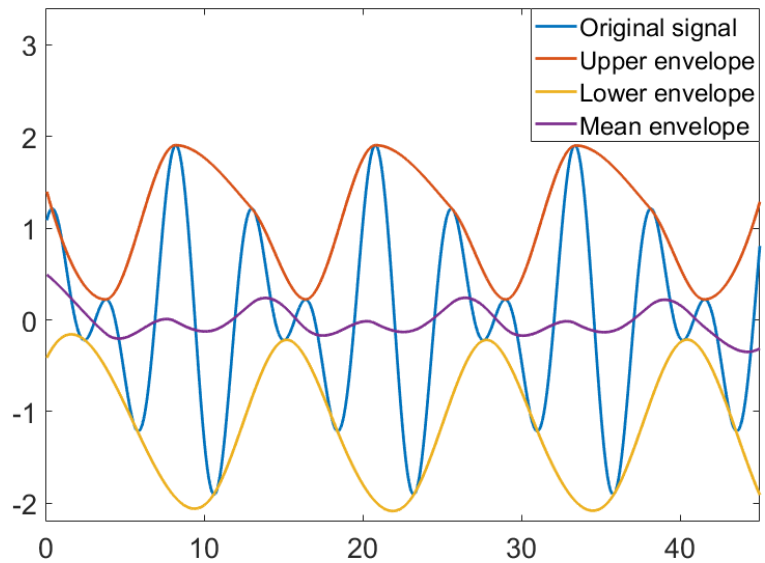

Figure 6. EMD schematic diagram

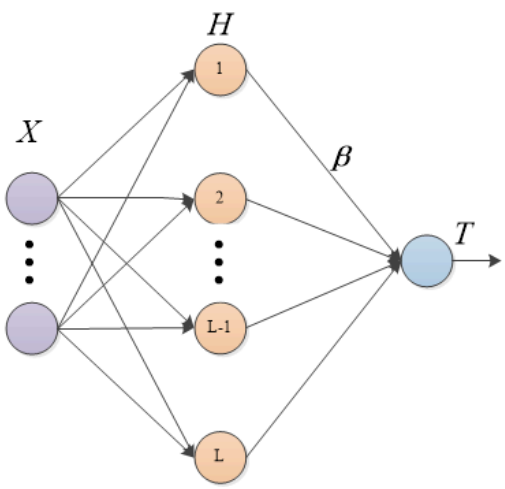

Figure 7. ELM structure diagram

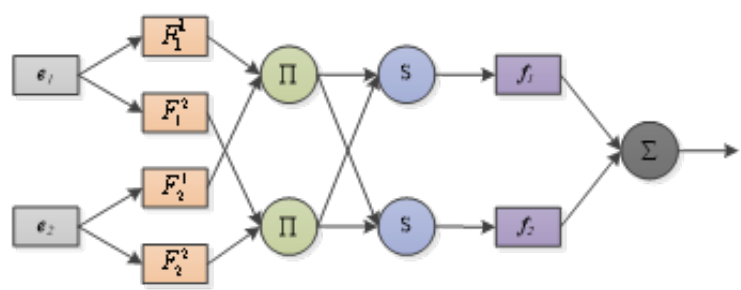

Figure 8. ANFIS structure diagram

Adaptive Network-based Fuzzy Inference System

Adaptive Network-based Fuzzy Inference System (ANFIS) is a kind of neural fuzzy inference system proposed by Jang J S R[32]. The ANFIS is widely used in various fields due to its advantages of convenience, high efficiency, and wide adaptability. If the ANFIS has two inputs e_1 and e_2, and one output y, the rule library has the following rules:

(1) If e_1 is A_1 and e_2 is B_1, then $y = p_1e_1 + q_1e_2 + z_1$;

(1) If e_1 is A_2 and e_2 is B_2, then $y = p_2e_1 + q_2e_2 + z_2$;

The typical ANFIS structure is shown in Figure 8, with a total of five layers. The first layer is the membership function layer of the input variables, which is responsible for the fuzzification of the input signals. Where, e_1 and e_2 are the inputs, F_i^j is the fuzzy set, representing the j-th rule of e_i, and the output result represents the degree to which the input e_i is subordinated to F_i^j. The second layer is used to release rules, the strength of the layer in the figure with $\prod$, said that it is multiplied by the input signal and its product output. The third layer is the normalization layer of rule strength, represented by N in the figure, which is responsible for calculating the normalized credibility of rule j of e_i. The fourth layer is the output layer of fuzzy rules, and each node of this layer is the adaptive node. The fifth layer is a fixed node that calculates the total output of all the input signals.

Method

The forecast flow chart of Empirical Mode Decomposition based Fuzzy Prediction Model (EMD-FPM) is shown in the Table 1. EMD-FPM is to decompose the concentration of air pollutants to obtain the variation trend of the concentration of air pollutants in different time scales, which is the IMF. Then, the IMF is trained with the ELM, and the ANFIS is used to fit the training results to get the final prediction value. The structure diagram is shown in Figure 9.

Algorithm 1 Empirical Mode Decomposition Fuzzy Forecast Model

Input: *Air data*;
Output: Predicted value y of AQI from day $t+1$ to day $t+c$
1: A=pretreat($Airdata$);
2: $IMF = (imf_1, imf_2, \ldots, imf_i, \ldots, imf_n)$=EMD($A$);
3: **for** i=1:n
4: x_i=win(imf_i);
5: ei=elmi(x_i);
6: **End**
7: $e = (e_1, e_2, \ldots, e_i, \ldots, e_n)$;
8: y=ANFIS(e)

Table 1. Empirical Mode Decomposition Fuzzy Forecast Model

In this paper, the EMD-FPM is described as follows, taking carbon monoxide (CO) as an example:

(1) CO concentration data of air pollutant is taken as input and normalized to obtain $CO(t)$, where t is the number of days;

(2) The EMD is used to process the data to obtain the corresponding IMF. The specific steps are as follows:

First, find all the maximum and minimum points of $CO(t)$, and then fit all the maximums and all the minimums through function fitting to obtain the upper envelope $up(t)$ and the lower envelope $low(t)$. $m_1(t)$ is the mean envelope of the original signal, which is the mean of the upper and lower envelopes:

$$m_1(t) = \frac{up(t) + low(t)}{2} \tag{8}$$

Subtract $m_1(t)$ from $CO(t)$, and get a new signal $h_1^1(t)$ with the low frequencies removed:

$$h_1^1(t) = CO(t) - m_1(t) \tag{9}$$

In fact, $CO(t)$ is complex and not regular. The calculated results generally do not meet the IMF conditions, so the above steps need to be repeated. Assume that $h_1^k(t)$ meets the conditions of the IMF after repeating the above steps for k times, then, the first order IMF component of the original signal is:

$$imf_1(t) = h_1^k(t) \tag{10}$$

The original signal $x(t)$ is subtracted by $imf_1(t)$ to obtain a new signal $r_1(t)$ with the high frequency component removed:

$$r_1(t) = x(t) - imf_1(t) \tag{11}$$

Repeat the above process for $r_1(t)$ to obtain a second IMF component $imf_2(t)$. And so on, until the end, IMF with n times of decomposition is obtained, $IMF = (imf_1, imf_2, \ldots, imf_i, \ldots, imf_n)$;

(3) The sliding window of size c+1 is used to slide IMF. The top c values of each sliding window are the input to the ELM, and the (c+1)-th value is the output of the ELM;

(4) The input and output obtained in step (3) are trained with the ELM to get elm_i. Then, the IMF corresponding to the air pollutant concentration is ELM, $ELM = (elm_1, elm_2, \ldots, elm_i, \ldots, elm_n)$;

(5) The predicted result is marked as e_i, then the predicted result corresponding to the ELM is $e = (e_1, e_2, \ldots, e_i, \ldots, e_n)$. The ANFIS is used for fitting, then the first layer output of the ANFIS is:

$$O_{ij}^1 = \mu_{F_i^j}(e_i) \tag{12}$$

Where, the membership function μ_F is determined by a number of parameters, which are called antecedent parameters;

The output of the second layer is as follows:

$$O_j^2 = \omega_j = \prod \mu_{F^j}(e_i) \tag{13}$$

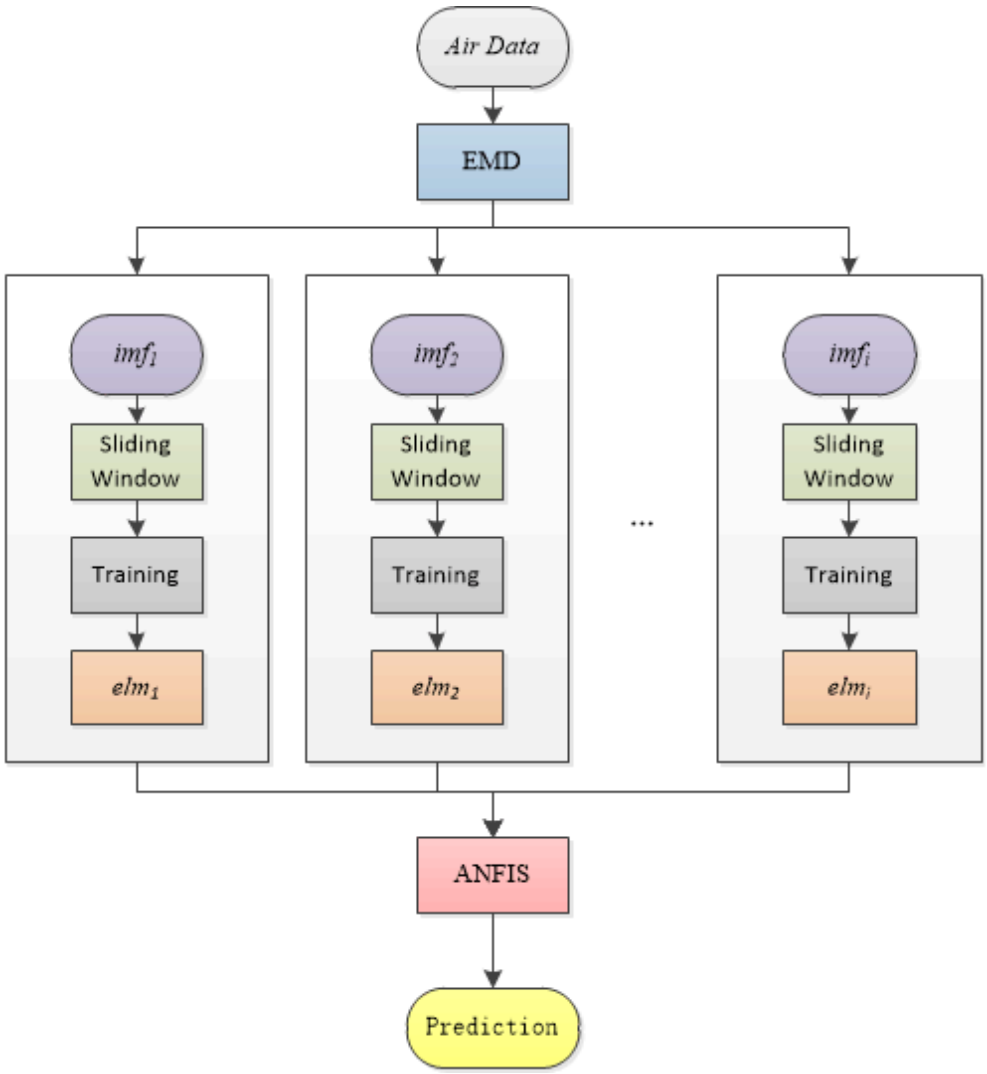

Figure 9. EMD-FPM structure diagram

The output of the third layer is as follows:

$$O_j^3 = \overline{\omega_j} = \frac{\omega_j}{\sum \omega_j} \tag{14}$$

The output of the fourth layer is as follows:

$$O_j^4 = \overline{\omega_j} f_j = \overline{\omega_j}(\sum p_i e_i + q_i) \tag{15}$$

The output of the fifth layer, that is, the final prediction result is as follows:

$$O^5 = \sum \overline{\omega_j} f_j \tag{16}$$

The final prediction results fit the predicted values of different time scales. The model has a good prediction effect for the data of a single city site.

References

1. Chen, Y. *et al.* Air quality data clustering using epls method. *Inf. Fusion* **36**, 225–232, DOI: https://doi.org/10.1016/j.inffus.2016.11.015 (2017).

2. Lee, H. J., Coull, B. A., Bell, M. L. & Koutrakis, P. Use of satellite-based aerosol optical depth and spatial clustering to predict ambient pm2. 5 concentrations. *Environ. research* **118**, 8–15 (2012).

3. Mahajan, S., Liu, H.-M., Tsai, T.-C. & Chen, L.-J. Improving the accuracy and efficiency of pm2.5 forecast service using cluster-based hybrid neural network model. *IEEE Access* **6**, 19193–19204, DOI: 10.1109/ACCESS.2018.2820164 (2018).

4. Xiao, Q., Chang, H. H., Geng, G. & Liu, Y. An ensemble machine-learning model to predict historical pm2.5 concentrations in china from satellite data. *Environ. Sci. & Technol.* **52**, 13260–13269, DOI: 10.1021/acs.est.8b02917 (2018). PMID: 30354085, https://doi.org/10.1021/acs.est.8b02917.

5. Wu, L., Gao, X., Xiao, Y., Liu, S. & Yang, Y. Using grey holt–winters model to predict the air quality index for cities in china. *Nat. Hazards* **88**, 1003–1012 (2017).

6. Xiong, P., Yin, Y., Shi, J. & Gao, H. Nonlinear multivariable gm (1, n) model based on interval grey number sequence. *J. Grey Syst.* **30** (2018).

7. Zalakeviciute, R., Bastidas, M., Buenaño, A. & Rybarczyk, Y. A traffic-based method to predict and map urban air quality. *Appl. Sci.* **10**, 2035 (2020).

8. Gu, K., Qiao, J. & Lin, W. Recurrent air quality predictor based on meteorology-and pollution-related factors. *IEEE Transactions on Ind. Informatics* **14**, 3946–3955 (2018).

9. Castelli, M., Clemente, F. M., Popovič, A., Silva, S. & Vanneschi, L. A machine learning approach to predict air quality in california. *Complexity* **2020** (2020).

10. Lyu, B., Zhang, Y. & Hu, Y. Improving pm2. 5 air quality model forecasts in china using a bias-correction framework. *Atmosphere* **8**, 147 (2017).

11. Liu, B. *et al.* A sequence-to-sequence air quality predictor based on the n-step recurrent prediction. *IEEE Access* **7**, 43331–43345 (2019).

12. Ray, R. *et al.* Prediction of benzene concentration of air in urban area using deep neural network. In *Proceedings of International Ethical Hacking Conference 2018*, 465–475 (Springer, 2019).

13. Qi, Z. *et al.* Deep air learning: Interpolation, prediction, and feature analysis of fine-grained air quality. *IEEE Transactions on Knowl. Data Eng.* **30**, 2285–2297 (2018).

14. Yang, Z. & Wang, J. A new air quality monitoring and early warning system: Air quality assessment and air pollutant concentration prediction. *Environ. research* **158**, 105–117 (2017).

15. Cheng, W., Shen, Y., Zhu, Y. & Huang, L. A neural attention model for urban air quality inference: Learning the weights of monitoring stations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32 (2018).

16. Yi, X., Zhang, J., Wang, Z., Li, T. & Zheng, Y. Deep distributed fusion network for air quality prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 965–973 (2018).

17. Li, X., Peng, L., Hu, Y., Shao, J. & Chi, T. Deep learning architecture for air quality predictions. *Environ. Sci. Pollut. Res.* **23**, 22408–22417 (2016).

18. Peng, H. *et al.* Evaluating hourly air quality forecasting in canada with nonlinear updatable machine learning methods. *Air Qual. Atmosphere & Heal.* **10**, 195–211 (2017).

19. Kabir, S., Islam, R. U., Hossain, M. S. & Andersson, K. An integrated approach of belief rule base and deep learning to predict air pollution. *Sensors* **20**, 1956 (2020).

20. [*et al.* Hourly pm2.5 concentration forecast using stacked autoencoder model with emphasis on seasonality. *J. Clean. Prod.* (2019).

21. Pardo, E. & Malpica, N. Air quality forecasting in madrid using long short-term memory networks. In *International Work-Conference on the Interplay Between Natural and Artificial Computation*, 232–239 (Springer, 2017).

22. Athira, V., Geetha, P., Vinayakumar, R. & Soman, K. Deepairnet: Applying recurrent networks for air quality prediction. *Procedia computer science* **132**, 1394–1403 (2018).

23. Soh, P.-W., Chang, J.-W. & Huang, J.-W. Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations. *Ieee Access* **6**, 38186–38199 (2018).

24. Zhou, Y., Chang, F.-J., Chang, L.-C., Kao, I.-F. & Wang, Y.-S. Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. *J. cleaner production* **209**, 134–145 (2019).

25. Lin, Y.-C., Lee, S.-J., Ouyang, C.-S. & Wu, C.-H. Air quality prediction by neuro-fuzzy modeling approach. *Appl. soft computing* **86**, 105898 (2020).

26. Yan, R. *et al.* Multi-hour and multi-site air quality index forecasting in beijing using cnn, lstm, cnn-lstm, and spatiotemporal clustering. *Expert. Syst. with Appl.* **169**, 114513 (2021).

27. Wang, J. & Song, G. A deep spatial-temporal ensemble model for air quality prediction. *Neurocomputing* **314**, 198–206 (2018).

28. Jin, W., Li, Z. J., Wei, L. S. & Zhen, H. The improvements of bp neural network learning algorithm. In *International Conference on Signal Processing* (2002).

29. Chow, T. W. S. & Leung, C. T. Neural network based short-term load forecasting using weather compensation. *IEEE Transactions on Power Syst.* **11**, 1736–1742 (1996).

30. Huang, N. E. *et al.* The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Math. Phys. & Eng. Sci.* **454**, 903–995 (1998).

31. Zhu, Q. Y. & Siew, C. K. Extreme learningmachine: A new learning scheme of feedforwardneuralnetworks. *Neurocomputing* **70**, 489–501 (2006).

32. Jang & J.-S., R. Anfis: adaptive-network-based fuzzy inference system. *IEEE Trans on Smc* **23**, 665–685 (1993).

Acknowledgements

This research was supported in part by the National Natural Science Foundation of China under grant No.61602202, by the Natural Science Foundation of Jiangsu Province under contracts BK20160428, by the Six Talent Peaks Project in Jiangsu Province under contract XYDXX-034 and by the major projects of natural science research in colleges and universities of Jiangsu Province.

Author contributions statement

B.L.C, G.C.Z and Y.Y.S conceived the ideas behind the paper. Q.X, M.J conducted the analysis of the datasets. B.L.C, G.C.Z and Y.T.Y carried out the statistical analysis. B.L.C, Y.Y.S, G.C.Z, Q.X, M.J and Y.T.Y analyzed the results. All authors wrote and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to GC.Z.